Extremal Theorems for Matrices

Doktori értekezés tézisei

Sali Attila
Rényi Alfréd Matematikai Kutató Intézet
H-1053 Budapest
Reáltanoda u. 13-15.
sali@renyi.hu

1. Bevezetés

Jelen disszertációban összegyűjtött eredmények vezér motívuma a mátrix formában megfogalmazható extremális kombinatorikai, illetve halmazrendszeres problémák. Halmazrendszerek természetes módon azonosíthatóak egyszerű $0-1$-mátrixokkal, amennyiben a sorok felelnek meg az alaphalmaz elemeinek, míg az oszlopok a halmazrendszerhez tartozó halmazok karakterisztikus vektorainak. Egy $m \times n$-es A mátrix *egyszerű*, ha bármely két oszlopa különböző. Ilyen értelmeben az extremális halmazrendszerek elméletének két alaptétele a következőképpen fogalmazható meg.

1. Tétel (Sperner, 1928). *Tegyük fel, hogy az egyszerű $m \times n$-es A mátrix bármely két oszlopában található* $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ *vagy* $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ *részmátrix. Ekkor $n \leq \binom{m}{\lfloor \frac{m}{2} \rfloor}$, egyenlőség esetén minden oszlop ugyanannyi 1-est tartalmaz.*

2. Tétel (Erdős-Ko-Rado, 1961). *Tegyük fel, hogy az egyszerű $m \times n$-es A mátrix bármely két oszlopában található* $\begin{bmatrix} 1 & 1 \end{bmatrix}$ *részmátrix és minden oszlopban k darab 1-es van. Ekkor ha $2k < m$, akkor $n \leq \binom{m-1}{k-1}$.*

A disszertáció három fő részből áll. Az első *tiltott részkonfigurációkkal*, vagy más néven *nyomokkal* foglalkozik. A második részben Sperner rendszerek *Vapnik-Chervonenkis dimenzióját* vizsgáljuk. Ezzel kapcsolatban bevezetjük a *rendezett szétzúzás* fogalmát. Az első két részben $0-1$-mátrixokkal foglalkozunk amelyek halmazrendszereket írnak le. A harmadik rész ezzel szemben relációs adatbázis modellek kombinatorikai problémáival foglalkozik. Egy relációs adatbázis legegyszerűbb modellje az a mátrix, melynek sorai az egyedi rekordoknak, oszlopai pedig az egyes tulajdonságoknak, azaz attribútumoknak felelnek meg. A különböző integritási feltételek az adatbázis mátrixokon érdekes extremális kombinatorikai problémákhoz vezetnek.

2. Tiltott részkonfigurációk

$0-1$-mátrixok tiltott részkonfigurációinak vizsgálata az extremális gráfelmélet hipergráfokra való kiterjesztésének is tekinthető, amellet, hogy az extremális halmazrendszerek elméletének része. Az egyszerű A mátrix egy az $\{1, 2, \ldots, m\}$ csúcshalmazú és n élű hipergráfot ír le, amennyiben a mátrix oszlopait az élek karakterisztikus vektorainak tekintjük. Azt mondjuk, hogy a $k \times l$-es F (nem feltétlenül egyszerű) $0-1$-mátrix az A mátrix *részkonfigurációja*, ha van A-nak olyan részmátrixa, amelyik F-ből sorok és oszlopok permutációjával kapható. Néha a részkonfigurációt *nyomnak* is nevezik és tekinthető a részgráf fogalom általánosításának hipergráfokra.

A 2. Fejezetben tárgyalt probléma a következő. Jelölje $\mathsf{forb}(m, F)$ a legkisebb olyan n értéket (m és F függvényében), amelyre igaz, hogy ha A egy egyszerű $m \times n$-es $0-1$-mátrix amelyik nem tartalmazza F-et részkonfigurációként, akkor $n \leq \mathsf{forb}(m, F)$. Az, hogy a definíció értelmes, és hogy $\mathsf{forb}(m, F) = O(m^k)$ Füredi egy észrevételéből [Für83] Sauer, Perles és Shelah, illetve Vapnik és Chervonenkis [Sau72, She72, VC71] tétele alapján következik.

A fejezet eredményei a [ABS09, AFFS05, AFS01, AGS97, ARS02, AS05, FS09] cikkekből származnak. A fő motiváció az [AS05]-ban leírt sejtés, ami a $\mathsf{forb}(m, F)$ nagyságrendjét adja meg és bizonyos értelemben az Erdős-Stone-Simonovits Tételre hasonlít. A 2.2.3. Sejtés azt mondja ki, hogy $\mathsf{forb}(m, F)$ nagyságrendi meghatározásához elegendő három alap mátrix típusból képzett direkt szozatokat vizsgálni. Ez a három típus az egységmátrix, annak $0-1$-komplementere, valamint az a felső háromszög mátrix, amlynek főátlójában és felette 1-esek vannak. Tegyük fel, hogy az $m_i \times n_i$-es A_i mátrixok egyszerűek $1 \leq i \leq t$ esetén. Ekkor t-szeres direkt szozat $A_1 \times A_2 \times \cdots \times A_t$ azt a $(\sum m_i) \times (\Pi n_i)$-es egyszerű mátrixot jelöli, melynek oszlopait úgy kapjuk, hogy az első m_1 sorra A_1 egy oszlopát tesszük, majd a következő m_2 sorra A_2 egy oszlopát, ... és így tovább, minden lehetséges kombinációban. A 2.2.3. Sejtés sejtés szerint $\mathsf{forb}(m, F) = \Theta(m^\ell)$ arra az ℓ természetes számra, melyre van olyan ℓ tényezős direkt szorzat, úgy hogy minden tényező a három alap mátrix egyike, és nincs F részkonfigurációja, viszont az alap mátrixok *bármely* $\ell + 1$ tényezős szorzata már tartalmazza F-et konfigurációként. A sejtés érdekessége, hogy $\mathsf{forb}(m, F)$ nagyságrendje mindig m *egész kitevős* hatványa.

A 2.3. alfejezetben a 2.2.3. Sejtést igazoljuk $k \times l$-es F-re $k \leq 3$ esetén. $k = 2$-re a 2.3.2. Tétel legérdekesebb esetében irányított gráfot definiálunk az F részkonfigurációt nem tartalmazó egyszerű A mátrix sorain, mint csúcshalmazon. F „hiánya" lefordítható ennek az irányított gráfnak a tulajdonságaira, amelynek segítségével kapjuk a felső becsléseket. Az alsó becslések a direkt szorzat konstrukcióból kaphatóak.

$k = 3$ esetben a 2.3.5. Tétel bizonyításában relációs adatbázisok funkcionális függőségeihez hasonló *implikációkat* vezetünk be. Ezen implikációk halmazából tudunk egy kvadratikus méretű fedő rendszert kiválasztani, ami a kvadratikus felső korlátok bizonyításának alapja.

A 2.2.3. Sejtés alapján két olyan maximális $k \times l$-es részkonfiguráció létezik, melynek tiltása a felső korlátot $\Theta(m^k)$-ról leviszi $O(m^{k-1})$-re. Ezek közül az egyiknek a helyességét bizonyítjuk a 2.3.3. alfejezetben, a 2.3.11. Tételben. A bizonyítás alapja az lemma, aminek segítségével az adott részkonfigurációt nem tartalmazó egyszerű A mátrixból el tudunk hagyni $O(m^{k-1})$ oszlopot úgy, hogy azok után már a Sauer, Perles és Shelah, illetve Vapnik és Chervonenkis tétel alkalmazható legyen rá. A lemma bizonyításának érdekessége, hogy elvezet Lovász egy 3-kritikus hipergráfokról szóló tételének [Lov76] erősítéséhez, illetve általánosításához. Ez a *partíció kritikus* illetve *rendezetten 3-kritikus* hipergráfok fogalmán alapszik, amelyeket a 2.5. alfejezetben vezetünk be.

Pontos eredmények teljes általánosságban a probléma természetéből adódóan nem várhatóak. Azonban, konkrét tiltott részkonfigurációkra teljesen pontos becslések adhatók. Ezeket gyűjtjük össze a 2.4. alfejezetben. Mivel a bizonyítások sokszor hosszadalmasak, ezért csak két 4×2-es konfigurációra vonatkozó eredményt írunk le részletesen. Ezek az [ABS09] cikkben fognak megjelenni. A 2.4.4. Tétel érdekessége a leszámlálási technika és az extremális rendszer karakterizációja. A 2.4.8. Tétel pedig rámutat a téma és a *kombinatorikus design elmélet* kapcsolatára. Azaz, az alsó korlát konstrukcióban a főtag együtthatója m növekedtével egymásba skatulyázott design-okkal javítható.

A 2.5. alfejezetben Toft és Lovász eredményeinek élesítését és általánosítását tárgyaljuk. Egy k-uniform hipergráf $\mathcal{H} = (V, \mathcal{E})$ *ℓ-kritikus*, ha nem $\ell - 1$-színezhető, de bármely csúcsát vagy élét elhagyva $\ell - 1$-színezhető hipergráfot kapunk. Toft bizonyította [Tof73], hogy rögzített $k, \ell > 3$ és $n \to \infty$, esetén létezik k-uniform ℓ-kritikus $\Omega(n^k)$ elű hipergráf n csúcson. Azonban, minden 3-kritikus k-uniform hipergráf élszáma $o(n^k)$. Toft kérdésére válaszolva Lovász bizonyította, hogy egy 3-kritikus k-uniform hipergráf élszáma legfeljebb $\binom{n}{k-1}$. A 2.5. alfejezet 2.5.5. Tételében rendezetten 3-kritikus hipergráfokra bizonyítjuk ugyanezt a felső korlátot, lineáris algebrai módszerekkel. Ezen kívül partíció kritikus hipergráfokra nagyságrendileg ugyanekkora felső korlátot adunk, valamint egy konstrukciót, melynek élszáma pontosan $\binom{n}{k-1}$. A felső és alsó korlát nagyságrendje $\Theta(n^{k-1})$, a különbségüké $\Theta(n^{k-3})$. A k-uniform $\mathcal{E} \subseteq \binom{[n]}{k}$ hipergráf az X n-elemű alaphalmazon *partíció kritikus* ha a következő feltételeket teljesíti. Az $\mathcal{E}$ élhalmazon adott egy sorbarendezés $E_1, E_2, \dots E_t$, valamint minden élhez elő van írva egy partíció $A_i \cup B_i = E_i$ ($A_i \cap B_i = \emptyset$), úgy hogy minden $i = 1, 2, \dots, t$-re létezik az alaphalmaznak egy partíciója $C_i \cup D_i = X$ ($C_i \cap D_i = \emptyset$) úgy, hogy $E_i \cap C_i = A_i$ és $E_i \cap D_i = B_i$, de sem $E_j \cap C_i \neq A_j$ sem $E_j \cap C_i \neq B_j$ $j < i$-re. Azaz, az alaphalmaz i-k partíciója az i-k élet az előírt módon vágja el, de semelyik korábbi élet sem az előírt módon vág szét. A hipergráf *rendezetten 3-kritikus*, ha minden i-re az előírt partíció $A_i = E_i$, $B_i = \emptyset$. Világos, hogy egy 3-kritikus hipergráf az rendezetten 3-kritikus is, és egy rendezetten 3-kritikus hipergráf az partíció kritikus is.

3. Antiláncok VC-dimenziója

A 3. Fejezetben amelynek kiinduló pontja Frankl [Fra89] sejtése, amelyik összekapcsolja az extremális halmazrendszerek elméletének két klasszikus eredményét, Sauer és Sperner tételeit, a [AS97, ARS02] cikkek eredményeit írjuk le. Mivel halmazrendszerek és egyszerű $0-1$ mátrixok azonosíthatóak, beszélhetünk halmazrendszerek részkonfigurációiról is, melyeket ebben a kontextusban *nyomnak* is szoktak nevezni. Jelölje K_k a $k \times 2^k$-as egyszerű $0-1$ mátrixot. Az $\mathcal{F} \subseteq 2^{[m]}$ halmazrendszer *Vapnik-Chervonenkis-dimenziója* (VC-dimenziója) az a legnagyobb k egész szám, amelyre $\mathcal{F}$-nek van K_k részkonfiurációja, illetve nyoma. Másképpen fogalmazva, az $\mathcal{F}$ halmazrenszer VC-dmenziója a legnagyobb olyan k egész szám, amelyre létezik az alaphalmaznak egy $|S| = k$ részhalmaza, melyre $|\{F \cap S \mid F \in \mathcal{F}\}| = 2^k$. Ekkor azt mondjuk, hogy $\mathcal{F}$ *szétzúzza* S-et. Frankl [Fra89] sejtése szerint ha $\mathcal{F}$ egy *antilánc*, amelyik nem zúz szét k vagy annál nagyobb elemszámú halmazt, akkor $\mathcal{F}| \leq \binom{m}{k-1}$. A 3.2. alfejezetben, a 3.2.4., 3.2.5. és 3.2.6. Tételekben Frankl sejtését bizonyítjuk be $k \leq 4$-re. A bizonyítás alapja indukció, és az, hogy $k \leq 3$-ra karakterizálni tudjuk az egyenlőség esetét.

A 3. Fejezetben tárgyalt fő fogalom a *rendezett szétzúzás* fogalma. Ez a klasszikus szétzúzás és a Bollobás és Radcliff [BLR89] által „fordított Sauer” egyenlőtlenségekhez bevezetett *strongly traced* fogalom közé esik, az alábbi értelemben. Jelölje $\mathsf{sh}(\mathcal{F})$ az $\mathcal{F}$ halmazrendszer által szétzúzott halmazok családját. (Ekkor $\mathsf{sh}(\mathcal{F})$ leszálló halmazrendszer és $\mathsf{sh}(\mathsf{sh}(\mathcal{F})) = \mathsf{sh}(\mathcal{F})$.) A rendezett szétzúzást S méretére vonatkozó indukcióval definiáljuk. $S = \emptyset$ esetén elegendő, ha $\mathcal{F}$ nem üres. Egyébként pedig azt mondjuk, hogy $\mathcal{F}$ rendezetten szétzúzza az $S = \{s_1, s_2, \dots, s_k\}$ halmazt ($s_1 < s_2 < \cdots < s_k$), ha létezik $\mathcal{F}$-nek $2^{|S|}$ eleme, melyek két halmazrendszerbe sorolhatóak, $\widetilde{\mathcal{F}_0}$-ba és $\widetilde{\mathcal{F}_1}$-be, úgy hogy $T = \{s_k + 1, s_k + 2, \dots, m\}$ esetén (T lehet üres halmaz) igaz az, hogy $T \cap C = T \cap D$ minden $C \in \widetilde{\mathcal{F}_0}$, $D \in \widetilde{\mathcal{F}_1}$, valamint $\{s_k\} \cap C = \emptyset, \{s_k\} \cap D = \{s_k\}$ minden $C \in \widetilde{\mathcal{F}_0}$, $D \in \widetilde{\mathcal{F}_1}$, továbbá $\widetilde{\mathcal{F}_0}$ és $\widetilde{\mathcal{F}_1}$ is külön-külön rendezetten szétzúzza $(S - \{s_k\})$-et. Jelölje $\mathsf{osh}(\mathcal{F})$ az $\mathcal{F}$ által rendezetten szétzúzott halmazok családját.Ekkor, hasonlóan $\mathsf{sh}(\mathcal{F})$-hez, igaz hogy $\mathsf{osh}(\mathcal{F})$ leszálló halmazrendszer és $\mathsf{osh}(\mathsf{osh}(\mathcal{F})) = \mathsf{osh}(\mathcal{F})$. Bollobás és Radcliff következőképpen definiálja a strongly traced fogalmat. $S \subseteq [m]$ strongly traced $\mathcal{F}$ szerint, ha létezik egy olyan $B \subseteq [m] - S$, amelyre $\{E \cap S : E \in \mathcal{F},\, E \cap ([m] - S) = B\} = 2^S$. A definíciók alapján világos, hogy $\mathsf{st}(\mathcal{F}) \subseteq \mathsf{osh}(\mathcal{F}) \subseteq \mathsf{sh}(\mathcal{F})$. Az $\mathsf{osh}(\mathcal{F})$ legfontosabb tulajdonsága, hogy $|\mathsf{osh}(\mathcal{F})| = |\mathcal{F}|$, amiből például Sauer, Perles és Shelah, Vapnik és Chervonenkis tétele azonnal következik.

A 3.3. alfejezetben először indukciót használva bizonyítjuk a 3.1.6. Tételt, ami az $\mathsf{osh}(\mathcal{F})$ előbb említett alap tulajdonságát mondja ki.

A 3.3.1. alfejezetben Frankl és Pach tételének [FP84], amelyik Frankl sejtése uniform halmazrendszerre, egy élesítését bizonyítjuk. A „nincs k méretű szétzúzott halmaz" feltételt helyettesítjük a „nincs k méretű rendezetten szétzúzott halmaz" feltétellel. A bizonyítás lényegi eleme az a karakterizáció, amit uniform halmazrenszerek által rendezetten szétzúzható halmazokra adunk a 3.3.3. Lemmában. Ennek igazi jelentősége nem csupán a tétel bizonyításában van, hanem az algebrai vonatkozásokban [ARS02, HR03b, HR03a, BRR06, HR06, BHR08] található. Egy halmazrendszer elemei természetesen azonosíthatóak monomialokkal, $F \subseteq [m]$-hez hozzárendelhető $x_F := \prod_{j \in F} x_j$, és viszont. Ha adott egy $\mathcal{F}$ halmazrendszer, akkor tekinthetjük azon m-változós polinomok $\mathcal{I}$ ideálját, melyek az $\mathcal{F}$-beli halmazok karakterisztikus vektorain 0 értéket vesznek fel. Ezen ideál standard monomjait lehet leírni a rendezett szétzúzás segítségével. A standard monomok kulcsszerepet játszanak a Gröbner bázisok elméletében. Jelölje $\mathsf{Sm}(\mathcal{F}) := \{F \subseteq [m] : x_F \in \mathsf{sm}(\mathcal{I})\}$, ahol $\mathsf{sm}(\mathcal{I})$ az $\mathcal{I}$ ideál standard monomjainak halmaza. Ekkor igaz, hogy $\mathsf{osh}(\mathcal{F}) = \mathsf{Sm}(\mathcal{F})$.

Mivel az uniform halmazrendszerrel rendezetten szétzúzható halmazok karakterizációja lehetőséget adott a Frankl sejtés megfelelő speciális esetének igazolására, ezért a következő lépés azon halmazok leírása, amelyeket antilánccal lehet rendezetten szétzúzni. A következő egyszerű numerikus karakterizációt adjuk meg a 3.3.2. alfejezet 3.3.5. Tételében.

3. Tétel (3.3.5. Tétel). *Legyen $S = \{s_1, s_2, \ldots, s_k\}$ az m-elemű alaphalmaz egy részhalmaza úgy, hogy $s_1 < s_2 < \cdots < s_k$. Létezik egy $\mathcal{A}$ antilánc, amelyre $S \in \mathsf{osh}(\mathcal{A})$, akkor és csak akkor, ha*

$$f(S) = \sum_{i=1}^{k} \frac{1}{2^{s_i - i}} < 1. \tag{1}$$

4. Adatbázis mátrixok

A 4. Fejezetben három különböző típusú problémával foglalkozunk, amelyek mindegyike relációs adatbázis modellek vizsgálata során kerül elő. Az új eredmények a [DKS92, DKS95, DKS98, sS98, ADKS00, AS07, SS08a, GOHKSS08, BS] cikkekből valóak.

A 4.3. alfejezet alapkérdése a következő. Tegyük fel, hogy $k \le n$, $p \le q < m$ pozitív egész számok és az $m \times n$-es M mátrix teljesíti az alábbi két tulajdonságot:

- Tetszőleges módon kiválasztva k különböző oszlopot, $c_1, c_2, \ldots, c_k$-t, létezik $q + 1$ sora M-nek úgy, hogy a különböző értkek száma ezekben a sorokban minden c_i $(1 \leq i \leq k-1)$ oszlopban legfeljebb p, azonban mind a $q+1$ érték a c_k oszlopban ezeken a sorokon különböző;

- Az előbbi feltétel már nem teljesül semmilyen $k+1$ különböző oszlop választása esetén sem.

A célunk m minimalizálása, rögzített n, p, q, k esetén.

A 4.4. alfejezetben egy adatbázisok motiválta kódelméleti problémát vizsgálunk. Egy q elemű ábécé feletti n hosszúságú kód *Armstrong(q, k, n)-kód*, ha a kódszavak minimális távolsága $n-k+1$, valamint tetszőleges $k-1$ koordináta pozícióhoz létezik két olyan kódszó, melyek ott egyeznek meg, azaz a minimális távolság „minden irányban” felvétetik.

A 4.5. alfejezetben egy diszkrepancia típusú eredményt bizonyítunk, amelyet adat olvasás optimalizálás motivál.

A 4.1. alfejezetben áttekintjük azokat a relációs adatbázis modellekhez kapcsolódó matematikai fogalmakat, amelyekre a 4. Fejezetben szükségünk lesz. Relációs adatbázis legegyszerűbb modellje egy mátrix, melynek oszlopai felelnek meg az attribútumoknak, azaz adat típusoknak, míg a sorai az egyes egyedek rekordjainak. Például egy munkahelyi adatbázis attribútumai lehetnek: Név, Anyja neve, Személyi szám, beosztás, Fizetés. Az adatbázis mátrix egy tipikus sora lehet (Nagy Jenő, Kiss Emeralda, 151543QW, portás, 97800). A matematikai modellben feltesszük, az általánosság korlátozása nélkül, hogy a mátrix elemei természetes számok. Egy adatbázishoz hozzátartoznak különböző *integritási feltételek* is. Ezek közül a legtöbbet használt és vizsgált fajta a *funkcionális függőség*. Az Y attribútum halmaz *funkcionális függ* az X attribútum halmaztól, ha egy rekord X-ben felvett értékei *egyértelműen* meghatározzák az Y-ban felvett értékeket. Azaz, ha a mátrix két sora megegyezik az X-beli pozíciókon, akkor megegyeznek az Y-belieken is. Relációs adatbázisok esetében megkülönböztetünk két fajta funkcionális függőséget. Az első az, amit tervezéskor előírnak, hogy teljesüljön, azaz tényleges integritási feltétel, a második fajta pedig az, ami az adatbázis pillanatnyi állapotában, az éppen aktuális adatbázis példányban teljesül, de nem következménye az előírt integritási feltételeknek.

Az $U \to V$ funkcionális függőség *logikai következménye* a Σ függőség halmaznak, jelölésben $\Sigma \models U \to V$, ha minden olyan adatbázis példányban, amiben Σ minden függősége teljesül, teljesül $U \to V$ is. A Σ (funcionális) függőség halmaz *Armstrong* példánya az r példány (adatbázis mátrix), ha $U \to V$ akkor és csak akkor teljesül r-ben, ha $\Sigma \models U \to V$. Funkcionális függőségi rendszerek Armstrong példányainak létezését Armstrong [Arm74] és Demetrovics [Dem79] bizonyították.

Egy függőségi rendszer minimális Armstrong példányának mérete a rendszer bonyolultságának egy mértéke. Adatbányászati szempontból tekintve, funkcionális függőségek keresésée esetén bizonyos függőségi rendszerek kizárhatóak a vizsgált példány mérete alapján. A 4.2. alfejezetben áttekintjük funkcionális függőségi rendszerek minimális Armstrong példányaival (reprezentációival) kapcsolatos eredményeket. Ezek igen bonyolult extremális kombinatorkai problémákhoz vezetnek. A felső becslésekhez használt konstrukciók sokszor design elmélet jellegűek. Az egyik esetben egy teljesen új vizsgálati rányt indítottak el, az *ortogonális kettős fedések* elméletét [BW90, GG87, Che92, GGM94, CD94, GMS95, Gro02].

A 4.3. alfejezetben funcionális függőségek egy általánosítását vezetjük be, és az azzal kapcsolatos kombinatorikai kérdéseket vizsgáljuk. A 4.3.1. Definíció szerint az $a \in \mathbf{R}$ attribútum (p, q)-függ az X attribútum halmaztól (jelölésben $X \xrightarrow{(p,q)} a$), ha az R relációnak (mátrixnak) nincs $q + 1$ olyan sora, melyek legfeljebb p különböző értéket tartalmaznak X-beli oszlopokban, azonban az a oszlopban felvett értékeik mind különbözőek. Az $(1, 1)$-függőség pontosan a funkcionális függőség. Ellentétben a funkcionális függőségi rendszerekkel, (p, q)-függőségeknek nem feltétlenül létezik Armstrong példányuk. Pontosabban fogalmazva, a következő a helyzet. A 4.2. alfejezetben leírjuk, funkcionális függőségek családjai ekvivalensek az attribútumok halmazán értelmezett lezárási operátorokkal, és ezen lezárások Armstrong példányait tekintjük. A 4.3. alfejezetben belátjuk, hogy a (p, q)-függőségek egy általánosabb fogalomhoz, a 4.3.2. Definícióban leírt *kiterjesztésekhez* vezetnek.

4. Definíció (4.3.2. Definíció, 4.3.3. Állítás). *Legyen Σ az $\mathbf{R}$ séma feletti (p, q)-függőségek egy családja. over the schema . Tegyük fel, hogy $1 \leq p \leq q$. A $\mathcal{J}_{\Sigma pq} \colon 2^{\mathbf{R}} \to 2^{\mathbf{R}}$ leképezést a*

$$\mathcal{J}_{\Sigma pq}(A) = \left\{ b \colon \Sigma \models A \xrightarrow{(p,q)} b \right\}. \tag{2}$$

formula definiálja. Ez a leképezés a következő két tulajdonsággal rendelkezik

$$\begin{array}{ll} (i) & A \subseteq \mathcal{J}_{\Sigma pq}(A) \\ (ii) & A \subseteq B \Longrightarrow \mathcal{J}_{\Sigma pq}(A) \subseteq \mathcal{J}_{\Sigma pq}(B). \end{array} \tag{3}$$

Az (i) és (ii) tulajdonsággal rendelkező leképezéseket kiterjesztéseknek *nevezzük*

A 4.3.1. alfejezetben elégséges feltételeket adunk arra, hogy egy kiterjesztésnek legyen Armstrong példánya, azaz (p, q)-függőséggel reprezentálható legyen, a 4.3.4. Tételben. A $p = q$ esetben a (p, p)-függőség által meghatározott kiterjesztés az lezárás is. Érdekes tehát vizsgálni, hogy milyen lezárásoknak létezik Armstrong példánya (p, p)-függőségek körében. Egy adott $\mathcal{L}$ lezárás *spektruma* $\mathsf{sp}\,(\mathcal{L})$ azon p természetes számokból áll, amelyekre $\mathcal{L}$-nek létezik Armstrong példánya (p, p)-függőségek körében. A 4.3.9. Tételben pontosan leírjuk az *uniform* lezárások spektrumát.

5. Tétel (4.3.9. Tétel). *Legyen $n \geq k^2\,(k-1)$ és jelölje $\mathcal{C}_n^k$ a k-uniform lezárást $\mathbf{R}$-n*

$$\mathcal{C}_n^k(X) = \begin{cases} X & \text{if } |X| < k \\ \mathbf{R} & \text{otherwise.} \end{cases} \tag{4}$$

Ekkor $\mathcal{C}_n^k$ spektruma $\mathsf{sp}\left(\mathcal{C}_n^k\right)$ a következő:

$$\mathsf{sp}\left(\mathcal{C}_n^k\right) = \{1, 2, \ldots, k-1\} \cup \{p \colon \exists s \in \mathbb{N}\; p + 1 - \left\lceil \frac{p+1}{s} \right\rceil = k - 1\}. \tag{5}$$

Az eredmény érdekessége, hogy a spektrumhoz tartozó „sporadikus” pontokat is sikerült megadni.

A 4.3.2. alfejezetben kiterjesztések és lezárások minimális Armstrong példányaival foglalkozunk, különféle p, q-függőségek esetében. Mivel a minimális reprezentáció már funkcionális függőségek, azaz $p = q = 1$ esetben is nehéz probléma, továbbá általános esetben maga az Armstrong példány létezésének kérdése is nehéz kérdés, ezért általános eredményeket nem várhatunk el. A 4.3.2. alfejezetben egy kiétellel csak uniform lezárásokkal foglalkozunk. A 4.3.21. Lemmában egy általános alsó korlátot adunk meg, ami a funkcionális függőségekre létező alsó korlát adaptációja. Az alfejezet fő eredményeiben konstrukciókkal bizonyítjuk, hogy a 4.3.21. Lemma alsó korlátja az adott esetekben nagyságrendileg helyes. A 4.3.22. Tételben véges projektív síkokat használunk a konstrukcióban, nem triviális módon. A 4.3.24. Tételben négy pontos eredményt gyűjtünk össze. Ezek közül kettő nagyságrendileg javít a 4.3.21. Lemma alsó korlátján. A bizonyítások közül csak az érdekesebbik kettőt vettük be a dolgozatba. A (ppn) esetben Lovász egy 1979-es tételét használjuk az alsó korlát bizonyítására, amelyik k-erdő hipergráfok maximális élszámát adja meg. Az (122) esetben a felső korlát érdekes. Ehhez egy n-elemű halmaz q-elemű részhalmazait kell úgy beosztanunk diszjunkt párokba, hogy ezek a párok egymás közt speciális metszet feltételt teljesítsenek (4.3.25. Tétel).

6. Tétel (4.3.25. Tétel). *Legyen $|X| = n$ és $2k > q$. X összes q-elemű részhalmazának családja partícionálható rendezetlen párokra (legfeljebb egy kivételével, ha $\binom{n}{q}$ páratlan), úgy, hogy a párosított q-elemű részhalmazok diszjunktak, továbbá ha A_1, B_1 és A_2, B_2 két ilyen pár, melyre $|A_1 \cap A_2| \geq k$, akkor $|B_1 \cap B_2| < k$, feltéve hogy $n > n_0(k, q)$.*

E tétel bizonyításához egy Dirac-típusú tételt mondunk ki speciális Hamilton-körök létezéséről (4.3.26. Tétel). A 4.3.25. Tétel érdekessége, hogy lehetővé teszi a diszjunkt k-elemű részhalmazok rendezetlen párjainak „terén" egy távolság megadását, és kódelméleti jellegű kérdesek vizsgálatát [EK01, BK01, BKL, KS04, Qui05, Qui09, DD06].

A 4.4. alfejezetben egy másik típusú kódelméleti kérdést tárgyalunk. Ezt korlátos értékkészletű attribútumok motiválják. Armstrong és Demetrovics eredményében, miszerint minden lezárásnak létezik Armstrong példánya funkcionális függőségek körében, szükséges feltételezés, hogy az egyes attribútumok értékkészlete tetszőlegesen nagy lehet. Azonban a *magasabbrendű adatmodell*, azaz *egymásba skatulyázott attribútumok* [HLS04, Sal04, SS06, SS08b] esetében a *számláló attribútumok* értékkészlete véges, valamint a valós életben is sok olyan helyzet fordul elő, amikor természetesen korlátos az egyes mezőkben felvehető értékek halmaza. Ilyen fordul elő például egy autó kölcsönző adatbázisnál, ahol az autó osztály besorolása csak a {mini, kompakt, alsó-közép, közép, felső, SUV, sport, minibusz} kategoriák egyike lehet.

A 4,4. alfejezet kiinduló pontja az a kérdés, hogy mlyen q, n, k értékekre létezik az n-elemű alaphalmazon k uniform lezárásnak Armstrong példánya, ha az attribútumok értékkészlete q elemű. Egy ilyen adatbázis mátrix sorai n hosszú, q elemű ábécé feletti kódszavaknak tekinthetőek. Ekkor semelyik két kódszó sem egyezhet meg k koordináta pozícióban, viszont bármely $k-1$ koordináta pozícióhoz léteznie kell két kódszónak, amelyek ott megegyeznek. Az ilyen kódokat nevezzük *Armstrong(q, k, n)-kódnak.* $f(q, k)$ jelöli azt a legnagyobb n értéket, amelyre Armstrong(q, k, n)-kód létezik. A 4.4.3. Tételben [GOHKSS08], alsó és felső becsléseket adunk $f(q, k)$-ra. Az egyik fő eredmény, hogy $q = 2$ esetben sikerül egy $c > 1$ konstans létezését bizonyítani melyre $\lfloor ck \rfloor \leq f(2, k)$. A 4.4.4. Állításban egy pontos és egy majdnem pontos értéket határozunk meg.

7. Állítás (4.4.4. Állítás). *$f(q,2) = \binom{q+1}{2}$ és $f(q,3) \le 3q-1$.*

Az utóbbi érdekessége, hogy a 4.2.8. Tétel, ami speciális típusú ortogonális kettős fedésekről szól és kombinatorikus design elméleti hátterű, ad a felső korlátnál csak eggyel kisebb alsó becslést, miszerint $f(q,3) \ge 3q-2$. Nagy k értékekre 4.4.5. Tételben [SS08a], sikerül a 4.4.4. Tétel alsó és felső korlátait megjavítani.

8. Tétel (4.4.5. Tétel). *$k > k_0(q)$ esetén*

$$\frac{\sqrt{q}}{e}k < f(q,k) < (q - \log q)k. \tag{6}$$

Az alsó korláthoz a véletlen konstrukciót adunk a Lovász Lokális Lemma használatával. A felső korláthoz az Armstrong(q,k,n)-kódot beágyazzuk az $n' = (q-1)n$-dimenziós euklideszi térbe mint egy szferikus kódot. Ehhez a lehetséges q szimbólumot egy $q-1$-dimenziós szabályos szimplex csúcsainak feleltetjük meg. A kód minimális távolsága meghatározza a szferikus kód minimális szögét. Ez Rankin egy tétele [Ran55] alapján felső becslést ad a szferikus kód pontszámára. Az Armstrong tulajdomság pedig, miszerint a minimális távolság minden irányban felvétetik, ad alsó becslést. A kettő összevetéséből kapjuk n-re a felső korlátot.

A 4.4.1. alfejezetben bináris Armstrong kódok konstrukcióit írjuk le. A 4.4.7. Állítás és a 4.4.8. Tétel [BS], bizonyításának alapja, hogy először egy kellően nagy minimális távolságú „váz-kódot” készítünk, majd az $n-k+1$-elemű koordináta pozíció halmazokat partícionáljuk úgy, hogy egy osztályba eső pozíció halmazok kellően távol legyenek egymástól. Az Armstrong kód a váz-kód szavaiból, valamint azoknak és a megfelelő pozíció halmazok karakterisztikus vektorainak összegeiből áll.

9. Tétel (4.4.8. Tétel). *Legyen $n-k = 2m$ vagy $n-k = 2m-1$ és $m > 1$. Ekkor Armstrong$(2,k,n)$-kód létezik, ha $n \ge 8m\log m$.*

A 4.5. alfejezetben egy diszkrepancia típusú eredményt tárgyalunk. Földrajzi, de egyéb adatbázisok is használják a 2-dimenziós képernyőt adatszervező eszközként. Azaz, a felhasználó kijelöli a képernyő egy területét, és az ahhoz tartozó adatokat kéri le. A modellt, amit használunk Bdel-Gafar és Abbadi [AGA97] vezette be. A feltételezés szerint az adatok párhuzamosan olvasható háttér tárolókon vannak, a minél gyorsabb adatolvasás érdekében a kijelölt képernyő területhez tartozó adatot minél több háttértárolón kell elosztani. A matematikai modellben feltesszük, hogy a felhasználó *téglalap* alakú területet jelöl ki. A képernyőt $n_1 \times n_2$ *csempére* osztjuk, egy csempéhez tartozó adatok egy háttér tárolón helyezkednek el. A felhasználó által kijelölt téglalapot két sarkának koordinátáival írhatjuk le $\mathcal{R} = \mathcal{R}[(i_1,j_1),(i_2,j_2)] = \{(i,j)\colon i_1 \le i \le i_2 \text{ és } j_1 \le j \le j_2\}$. Minden (i,j) csempéhez egy $f(i,j)$, 1 és m közé eső, egész számot rendelünk ami azt mondja meg, hogy a csempe adata melyik tárolón van. Egy ilyen hozzárendelés akkor jó, ha minden előforduló téglalapra, a benne legtöbbször, illetve legkevesebbszer szereplő tároló szám előfordulásának számai közt a különbség a kicsi.

Definàljuk egy $f(i,j)$ hozzárendelés diszkrepancáját, majd ezt használva az m szám diszkrepanciáját. Latin négyzeteket használunk optimális hozzárendelés megadásához. A konstrukció indukción alapul, latin négyzetek direkt szozatát használjuk. Továbbá szükségünk van egyfajta „összeadás” lehetőségére is latin négyzetek között. Ehhez definiáljuk egy transzverzális diszkrepanciáját, majd ezt használva tudunk $n \times n$-es latin négyzetről $n + 1 \times n + 1$-esre áttérni. Az alfejezet két fő tétele a 4.5.4. és 4.5.5. tételek, amelyek logaritmikus diszkrepancájú hozzárendelést adunk meg és bizonyítjuk, hogy latin négyzet típusú hozzárendeléssel ez az lehető legjobb. Az alfejezet anyaga a [ADKS00] konferencia cikken és [AS07] folyóirat cikken alapszik.

5. Köszönetnyilvánítás

Sokaknak tartozom köszönettel, mert tanítottak, segítettek pályám során. Szakmai pályafutásom elindítója, évtizedeken át támogatója és mind a mai napig meghatározója Katona Gyula, akinek nemcsak a matematikai támogatásért, hanem barátságáért is köszönettel tartozom. Kandidátusi disszertációm témavezetője Füredi Zoltán volt, akitől azután is sok szakmai segítséget kaptam és hatékony módszereket tanultam. Hálás vagyok Demetrovics Jánosnak, aki adatbázis elméleti cikkeim legtöbbjének társszerzője. Külön köszönet illeti Simonyi Gábort, akivel élmény volt együtt dolgozni gráfelméleti problémákon. Sok-sok figyelmet és segítséget köszönök Recski Andrásnak, Simonovits Miklósnak és T. Sós Verának. Vendégszeretetükért és barátságukért illeti köszönet Hamburger Pétert és Székely Lászlót. Végül, de nem utolsó sorban, rendkívüli hálás vagyok Richard Anstee-nek hogy bevezetett a tiltott részkonfigurációk és a rendezett szétzúzás elméletébe.

Természetesen köszönettel és hálával tartozom családomnak is. Feleségem, Kovács Ildi szakadatlan és feltétlen hite bennem segített, hogy a szakmai munkára tudjak koncentrálni. Szüleimnek köszönöm, hogy életem nehéz pillanataiban is mindig mellettem álltak. Fiaim a legtöbbel ajándékoztak meg, amit egy apa kaphat, a barátságukkal. Köszönöm nekik az együtt sportolás élményét.

A disszertáció alpjául szolgáló cikkek

[ABS09] R.P. Anstee, F. Barekat, and A. Sali, *Small Forbidden Configurations V: Exact bounds for* 4×2 *cases*, Studia Sci. Math. Hun. (2009), 20 pp., to appear.

[AFFS05] R.P. Anstee, B. Fleming, Z. Füredi, and A. Sali, *Color critical hypergraphs and forbidden configurations*, Discrete Mathematics and Theoretical Computer Science Proceedings (S. Felsner, ed.), vol. AE, 2005, pp. 117–122.

[AFS01] R. P. Anstee, R. Ferguson, and A. Sali, *Small Forbidden Configurations II*, Electronic J. Combin. **8** (2001), R4 (25 pp).

[AGS97] R. P. Anstee, J. R. Griggs, and A. Sali, *Small forbidden configurations*, Graphs and Combinatorics **13** (1997), 97–118.

[ARS02] R. P. Anstee, L. Rónyai, and A. Sali, *Shattering news*, Graphs and Combin. **18** (2002), 59–73.

[AS97] R.P. Anstee and A. Sali, *Sperner families of bounded VC-dimension*, Discrete Math. **175** (1997), 13–21.

[AS05] R. P. Anstee and A. Sali, *Small Forbidden Configurations IV*, Combinatorica **25** (2005), 503–518.

[AS07] R. P. Anstee and A. Sali, *Latin squares and low discrepancy allocation of two-dimensional data*, European J. of Combinatorics **28** (2007), 2116–2124.

[BS] A. Blokhuis and A. Sali, Paper in preparation.

[DKS92] J. Demetrovics, G.O.H. Katona, and A. Sali, *The characterization of branching dependencies*, Discrete Applied Mathematics **40** (1992), 139–153.

[DKS95] J. Demetrovics, G.O.H. Katona, and A. Sali, *Representations of branching dependencies*, Acta Sci. Math. (Szeged) **60** (1995), 213–223.

[DKS98] J. Demetrovics, G.O.H. Katona, and A. Sali, *Design type problems motivated by database theory*, Journal of Statistical Planning and Inference **72** (1998), 149–164.

[FS09] Z. Füredi and A. Sali, *Partition critical hypergraphs*, in preparation, 2009.

[GOHKSS08] Gyula O. H. Katona, Attila Sali, and Klaus-Dieter Schewe, *Codes that attain minimum distance in all possible directions*, Central Eruopean J. of Math. **6** (2008), 1–11.

[Sal04] Attila Sali, *Minimal keys in higher-order datamodels*, Foundations of Information and Knowledge Systems (Dietmar Seipel and José María Turull Torres, eds.), Springer LNCS, vol. 2942, Springer Verlag, 2004.

[sS98] Attila Sali sr. and Attila Sali, *Generalized dependencies in relational databases*, Acta Cybernetica (Szeged) **13** (1998), 431–438.

[SS08a] A Sali and L. Székely, *On the existence of armstrong instances with bounded domains*, Lecture Notes in Computer Science (S. Hartmann and G. Kern-Isberner, eds.), vol. 4932, Springer, 2008, pp. 151–157.

Hivatkozások

[ABS09] R.P. Anstee, F. Barekat, and A. Sali, *Small Forbidden Configurations V: Exact bounds for* 4×2 *cases*, Studia Sci. Math. Hun. (2009), 20 pp., to appear.

[ADKS00] R.P. Anstee, J. Demetrovics, G.O.H. Katona, and A. Sali, *Low discrepancy allocation of two dimensional data*, Lecture Notes in Computer Science (K.-D. Schewe and B. Thalheim, eds.), vol. 1762, Springer, 2000, pp. 1–12.

[AFFS05] R.P. Anstee, B. Fleming, Z. Füredi, and A. Sali, *Color critical hypergraphs and forbidden configurations*, Discrete Mathematics and Theoretical Computer Science Proceedings (S. Felsner, ed.), vol. AE, 2005, pp. 117–122.

[AFS01] R. P. Anstee, R. Ferguson, and A. Sali, *Small Forbidden Configurations II*, Electronic J. Combin. **8** (2001), R4 (25 pp).

[AGA97] K. A. S. Abdel-Ghaffar and A. El Abbadi, *Optimal allocation of two dimensional data*, Lecture Notes in Computer Science (F. Afrati and Ph. Kolaitis, eds.), vol. 1186, Springer, 1997, pp. 409–418.

[AGS97] R. P. Anstee, J. R. Griggs, and A. Sali, *Small forbidden configurations*, Graphs and Combinatorics **13** (1997), 97–118.

[Arm74] W. W. Armstrong, *Dependency structures of database relationships*, Information Processing (1974), 580–583.

[ARS02] R. P. Anstee, L. Rónyai, and A. Sali, *Shattering news*, Graphs and Combin. **18** (2002), 59–73.

[AS97] R.P. Anstee and A. Sali, *Sperner families of bounded VC-dimension*, Discrete Math. **175** (1997), 13–21.

[AS05] R. P. Anstee and A. Sali, *Small Forbidden Configurations IV*, Combinatorica **25** (2005), 503–518.

[AS07] ———, *Latin squares and low discrepancy allocation of two-dimensional data*, European J. of Combinatorics **28** (2007), 2116–2124.

[BHR08] B.Felszeghy, G. Hegedűs, and L. Rónyai, *Algebraic properties of modulo q complete ℓ-wide families*, 2008, Published online by Cambridge University Press 09 Dec 2008 doi:10.1017/S0963548308009619.

[BK01] G. Brightwell and G.O.H. Katona, *A new type of coding theorem*, Studia Scientiarum Mathematicarum Hungarica **38** (2001), 139–147.

[BKL] B. Bollobás, G.O.H. Katona, and I. Leader, *A coding problem for pairs of subsets*, Manuscript under preparation.

[BLR89] B. Bollobás, I. Leader, and A.J. Radcliffe, *Reverse Kleitman inequalities*, Proc. London Math. Soc. **58** (1989), 153–168.

[BRR06] B.Felszeghy, B. Ráth, and L. Rónyai, *The lex game and some applications*, Journal of Symbolic Computation **41** (2006), 663–681.

[BS] A. Blokhuis and A. Sali, Paper in preparation.

[BW90] F.E. Bennett and Lisheng Wu, *On mimimum matrix representation of closure operations*, Discrete Appl. Math. **26** (1990), 25–40.

[CD94] M.S. Chung and D.B.West, *The p-intersection number of a complete bipartite graph and orthogonal double coverings of a clique*, Combinatorica **14** (1994), 453–461.

[Che92] Yeow Meng Chee, *Design-theoretic problems in perfectly* $(n-3)$*-error-correcting databases*, preprint, 1992.

[DD06] Michel-Marie Deza and Elena Deza, *Dictionary of distances*, Elsevier, 2006.

[Dem79] J. Demetrovics, *On the equivalence of candidate keys with Sperner systems*, Acta Cybernetica **4** (1979), 247–252.

[DKS92] J. Demetrovics, G.O.H. Katona, and A. Sali, *The characterization of branching dependencies*, Discrete Applied Mathematics **40** (1992), 139–153.

[DKS95] ______, *Representations of branching dependencies*, Acta Sci. Math. (Szeged) **60** (1995), 213–223.

[DKS98] ______, *Design type problems motivated by database theory*, Journal of Statistical Planning and Inference **72** (1998), 149–164.

[EK01] H. Enomoto and G.O.H. Katona, *Pairs of disjoint q-element subsets far from each other*, Electr. J. Comb. **8** (2001), #R7.

[FP84] P. Frankl and J. Pach, *On disjointly representable sets*, Combinatorica **4** (1984), 39–45.

[Fra89] P. Frankl, *Traces of antichains*, Graphs and Combin. **5** (1989), 295–299.

[FS09] Z. Füredi and A. Sali, *Partition critical hypergraphs*, in preparation, 2009.

[Für83] Z. Füredi, 1983, Personal communication to R.P. Anstee.

[GG87] H.-D.O.F. Gronau and B. Ganter, *On two conjectures of Demetrovics, Füredi and Katona concerning partitions*, Discrete Mathematics **88** (1987), 149–155.

[GGM94] B. Ganter, H.-D. O. F. Gronau, and R. C. Mullin, *On orthogonal double covers of* K_n, Ars Combinatoria **37** (1994), 209–221.

[GMS95] H.-D.O.F. Gronau, R.C. Mullin, and P.J. Schellenberg, *On orthogonal double covers of K_n and a conjecture of Chung and West*, J. of Combinatorial Designs **3** (1995), 213–231.

[GOHKSS08] Gyula O. H. Katona, Attila Sali, and Klaus-Dieter Schewe, *Codes that attain minimum distance in all possible directions*, Central Eruopean J. of Math. **6** (2008), 1–11.

[Gro02] H.-D.O.F. Gronau, *On orthogonal double covers of graphs*, Designs, Codes and Cryptography **27** (2002), 49–91.

[HLS04] Sven Hartmann, Sebastian Link, and Klaus-Dieter Schewe, *Weak functional dependencies in higher-order datamodels*, Foundations of Information and Knowledge Systems (Dietmar Seipel and José María Turull Torres, eds.), Springer LNCS, vol. 2942, Springer Verlag, 2004.

[HR03a] G. Hegedűs and L. Rónyai, *Gröbner bases for complete uniform families*, Journal of Algebraic Combinatorics **17** (2003), 171–180.

[HR03b] ______, *Standard monomials for q-uniform families and a conjecture of Babai and Frankl*, Central European Journal of Mathematics **1** (2003), 198–207.

[HR06] ______, *Standard monomials for partitions*, Acta Mathematica Hungarica **111** (2006), 193–212.

[KS04] G.O.H. Katona and A. Sali, *New type of coding problem motivated by data base theory*, Discr. Appl. Math. **144** (2004), 140–148.

[Lov76] L. Lovász, *Chromatic number of hypergraphs and linear algebra*, Studia Sci. Math. Hung. **11** (1976), 113–114.

[Qui05] J. Quistorff, *New Upper Bounds on Enomoto-Katona's Coding Type Problem*, Studia Sci. Math. Hungar. **42** (2005), 61–72.

[Qui09] ______, *Combinatorial problems in the Enomoto-Katona space*, to appear in Studia Sci. Math. Hungar., 2009.

[Ran55] R.A. Rankin, *The closest packing of spherical caps in n dimensions*, Proceedings of the Glasgow Mathematical Society **2** (1955), 145–146.

[Sal04] Attila Sali, *Minimal keys in higher-order datamodels*, Foundations of Information and Knowledge Systems (Dietmar Seipel and José María Turull Torres, eds.), Springer LNCS, vol. 2942, Springer Verlag, 2004.

[Sau72] N. Sauer, *On the density of families of sets*, J. Combin. Th. Ser A **13** (1972), 145–147.

[She72] S. Shelah, *A combinatorial problem: Stability and order for models and theories in infinitary languages*, Pac. J. Math. **4** (1972), 247–261.

[sS98] Attila Sali sr. and Attila Sali, *Generalized dependencies in relational databases*, Acta Cybernetica (Szeged) **13** (1998), 431–438.

[SS06] Attila Sali and Klaus-Dieter Schewe, *Counter-free keys and functional dependencies in higher-order datamodels*, Fundamenta Informaticae **70** (2006), 277–301.

[SS08a] A Sali and L. Székely, *On the existence of armstrong instances with bounded domains*, Lecture Notes in Computer Science (S. Hartmann and G. Kern-Isberner, eds.), vol. 4932, Springer, 2008, pp. 151–157.

[SS08b] Attila Sali and Klaus-Dieter Schewe, *Keys and Armstrong databases in trees with restructuring*, Acta Cybernetica **18** (2008), 529–556.

[Tof73] B. Toft, *On colour-critical hypergraphs*, Colloquia Mathematica Societatis János Bolyai, Infinite and Finite Sets, Keszthely (Hungary), János Bolyai Mathematical Society, 1973, pp. 1445–1457.

[VC71] V.N. Vapnik and A.Ya. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Th. Prob. and Applics. **16** (1971), 264–280.